

Claims

We claim:

1. A computerized data mining system, comprising:

a data exploration system for receiving and analyzing user data to provide statistical information about the user data;

a customized model system for generating and ranking customized data mining models, and for executing a selected customized data mining model on the user data, wherein the customized data mining models are iteratively generated in parallel based on permutations of at least one of the user data, business parameters and a set of model generation algorithms; and

an existing model system for selecting at least one existing data mining model from a library of existing data mining models, and for executing the selected at least one existing data mining model in parallel on the user data.

2. The computerized data mining system of claim 1, further comprising:

a data submission system for submitting the user data; and

a parameter designation system for designating the business parameters.

3. The computerized data mining system of claim 1, wherein the customized model system comprises:

a model generation system for iteratively generating the customized data mining models in parallel based on the permutations of at least one of the user data, the business parameters and the set of model generation algorithms;

a model ranking system for ranking the customized data mining models based on the business parameters, for identifying a predetermined quantity of the ranked customized data mining models, and for providing comparative data corresponding to the predetermined quantity of the ranked customized data mining models;

a customized model selection system for selecting at least one customized mining model from the predetermined quantity; and

a customized model execution system for executing the selected at least one customized data mining model on the user data.

4. The computerized data mining system of claim 1, wherein the existing model system comprises:

a model library system for assembling the library of existing data mining models based on the business parameters, and for displaying the library of existing data mining models and comparative data corresponding to the library of existing data models;

an existing model selection system for selecting the at least one existing data mining model from the library of existing data mining models;

an existing model execution system for executing the at least one existing data mining model on the user data in parallel; and

a existing model comparison system for comparing results of the execution of the at least one existing data mining model.

5. The computerized data mining system of claim 1, wherein the business parameters comprise a business taxonomy, a set of business problems and a set of model goals.

6. The computerized data mining system of claim 1, wherein the statistical information comprises data relationships, data outliers, invalid data values and standard deviations.

7. The computerized data mining system of claim 1, wherein the computerized data mining system is implemented in a network environment.

8. A computerized system for generating and executing customized data mining models, comprising:

a model generation system for iteratively generating the customized data mining models in parallel based on the permutations of at least one of user data, business parameters and a set of model generation algorithms;

a model ranking system for ranking the customized data mining models based on the business parameters, for identifying a predetermined quantity of the ranked customized data mining models, and for providing comparative data corresponding to the predetermined quantity of the ranked customized data mining models;

a customized model selection system for selecting at least one customized mining model of the customized data mining models; and

a customized model execution system for executing the selected at least one customized data mining model on the user data.

9. The computerized system of claim 8, wherein the computerized data mining system is implemented in a network environment.

10. The computerized system of claim 8, wherein the business parameters comprise a business taxonomy, a set of business problems and a set of model goals.

11. A computerized system for selecting and executing existing data mining models:

 a model library system for assembling a library of existing data mining models based on a business parameters, and for displaying the library of existing data mining models and comparative data corresponding to the library of existing data models;

 an existing model selection system for selecting at least one existing data mining model from the library of existing data mining models;

 an existing model execution system for executing the at least one existing data mining model on the user data in parallel; and

 an existing model comparison system for comparing results of the execution of the at least one existing data mining model.

12. The computerized system of claim 11, wherein the at least one existing data mining model is executed on the user data in a grid environment.

13. The computerized system of claim 11, wherein the business parameters comprise a business taxonomy, a set of business problems and a set of model goals.

14. A computer-implemented method for generating customized data mining models, comprising:

providing user data and business parameters;

iteratively generating a plurality of customized data mining models in parallel based on permutations of at least one of the user data, the business parameters and a set of model generation algorithms;

ranking the plurality of customized data mining models based on the business parameters;

selecting at least one customized data mining model from the ranked plurality of customized data mining models; and

executing the selected at least one customized data mining model on the user data.

15. The computer-implemented method of claim 14, further comprising:

identifying a predetermined quantity of the ranked plurality of customized data mining models; and

providing comparative data corresponding to the predetermined quantity of the ranked plurality of customized data mining models, prior to the selecting step.

16. The computer-implemented method of claim 14, wherein the step of iteratively generating comprises:

forming multiple permutations of at least one of the user data, the business parameters and the set of model generation algorithms; and

iteratively generating a plurality of customized data mining models in a grid environment based on the multiple permutations.

17. The computer-implemented method of claim 14, wherein the statistical information comprises data relationships, data outliners, invalid data values and standard deviations.

18. The computer-implemented method of claim 14, wherein the business parameters comprise a business taxonomy, a set of business problems and a set of model goals.

19. A computer-implemented method for selecting existing data mining models, comprising:

providing user data and business parameters;

assembling a library of existing data mining models based on the business parameters;

displaying the library of existing data mining models and comparative data corresponding to the library of data mining models;

selecting at least one existing data mining model from the library of existing data mining models;

executing the at least one existing data mining model on the user data in parallel;

and

comparing results of the execution of the at least one existing data mining model.

20. The computer-implemented method of claim 19, wherein the business parameters comprise a business taxonomy, a set of business problems and a set of model goals.

21. A data mining computer program product stored on a recordable medium, which when executed, comprises:

- program code for receiving and analyzing user data to provide statistical information about the user data;

- program code for generating and ranking customized data mining models, and for executing a selected customized data mining model on the user data, wherein the customized data mining models are iteratively generated in parallel based on permutations of at least one of the user data, business parameters and a set of model generation algorithms; and

- program code for selecting at least one existing data mining model from a library of existing data mining models, and for executing the selected at least one existing data mining model in parallel on the user data

22. The data mining computer program product of claim 21, further comprising:

- program code for submitting the user data; and

- program code for designating the business parameters.

23. The data mining computer program product of claim 21, wherein the program code for generating and ranking comprises:

program code for iteratively generating the customized data mining models in parallel based on the permutations of at least one of the user data, the business parameters and the set of model generation algorithms;

program code for ranking the customized data mining models based on the business parameters, for identifying a predetermined quantity of the ranked customized data mining models, and for providing comparative data corresponding to the predetermined quantity of the ranked customized data mining models;

program code for selecting at least one customized mining model from the predetermined quantity; and

program code for executing the selected at least one customized data mining model on the user data.

24. The data mining computer program product of claim 21, wherein the program code for selecting comprises:

program code for assembling the library of existing data mining models based on the business parameters, and for displaying the library of existing data mining models and comparative data corresponding to the library of existing data models;

program code for selecting the at least one existing data mining model from the library of existing data mining models;

program code for executing the at least one existing data mining model on the user data in parallel; and

program code for comparing results of the execution of the at least one existing data mining model.

25. The data mining computer program product of claim 21, wherein the business parameters comprise a business taxonomy, a set of business problems and a set of model goals.

26. The data mining computer program product of claim 21, wherein the statistical information comprises data relationships, data outliers, invalid data values and standard deviations.

27. The data mining computer program product of claim 21, wherein the program product is implemented in a network environment.

28. A program product stored on a recordable medium for generating and executing customized data mining models, which when executed comprises:

program code for iteratively generating the customized data mining models in parallel based on the permutations of at least one of user data, business parameters and a set of model generation algorithms;

program code for ranking the customized data mining models based on the business parameters, for identifying a predetermined quantity of the ranked customized data mining models, and for providing comparative data corresponding to the predetermined quantity of the ranked customized data mining models;

program code for selecting at least one customized mining model from the predetermined quantity; and

program code for executing the selected at least one customized data mining models on the user data.

29. The program product of claim 28, wherein the program product is implemented in a network environment.

30. A program product stored on a recordable medium for selecting and executing existing data mining models, which when executed comprises:

program code for assembling a library of existing data mining models based on business parameters, and for displaying the library of existing data mining models and comparative data corresponding to the library of existing data models;

program code for selecting at least one existing data mining model from the library of existing data mining models;

program code for executing the at least one existing data mining model on the user data in parallel; and

program code for comparing results of the execution of the at least one existing data mining model.

31. The program product of claim 30, wherein the at least one existing data mining model is executed on the user data in a grid environment.

32. The program product of claim 30, wherein the business parameters comprise a business taxonomy, a set of business problems and a set of model goals.